

# 1 Examples

## 1.1 Detecting signal from noise

Reference: D.S.Sivia: Data analysis, a Bayesian tutorial, 2nd ed. p.35-42.

Assume we have measurements (counts)  $y_i$  taken at points  $x_i$ . The simple model is Poisson with parameter  $e_i$ . This expected value is assumed to be a combination of background noise and the signal so that:

$$e_i = 100(ae^{-0.5(x_i-x_0)^2/s^2} + b)$$

And we assume the location and the width of the signal to be known, specified by constants  $x_0 = 0, s = 1$ . The unknown parameters to be estimated are the signal amplitude  $a$  and the level of noise  $b$ . Generate some data  $y_i$  by choosing values for  $a, b$  and  $x_i$ . Study the 2D-posterior distribution of  $(a, b)$ , based on uniform priors over sufficient range, e.g.  $U(0, 1000)$ . Try with different number of data points and study how the posterior behaves.

```
model{
x0 <- 0; s <- 1; v<- s*s # assumed known constants
#a <- 1; b <- 2 # true values
a ~ dunif(0,1000); b ~dunif(0,1000) # priors
for(i in 1:L+1){
x[i] <- -5 + 10*(i-1)/L # measurement points
y[i] ~dpois(e[i])
e[i] <- 100*(a *exp(-0.5*pow(x[i]-x0,2)/v) +b) # expected value (signal+noise)
}
}
list(L=2,y = c(212,299,194))
list(L=10,y = c(174,200,203,225,275,308,269,223,170,192,197))
list(L=100,
y = c(
225,201,189,206,222,
201,209,202,209,172,
182,213,176,215,213,
171,195,211,216,203,
175,192,197,218,203,
175,219,186,209,229,
204,238,222,243,233,
248,272,253,222,261,
243,269,252,257,301,
289,297,267,303,281,
306,293,291,280,288,
293,288,262,276,269,
265,241,271,269,243,
238,231,238,212,190,
215,209,194,220,208,
```

199,218,228,200,207,  
 201,206,180,196,205,  
 215,205,204,179,207,  
 193,208,202,206,195,  
 215,182,213,215,183,  
 199))

## 1.2 Modeling sensitivity and prevalence

Assume  $n$  populations in which the disease prevalences are  $p_1, \dots, p_n$ . A sample of  $K$  individuals is collected from each and tested. The test has sensitivity  $p_s$ . Generate data by randomly generating  $p_i \sim U(0, 1)$  and  $x_i \sim \text{Bin}(p_i p_s, K)$  with a chosen  $p_s$  and  $K$ . Estimate  $p_1, \dots, p_n$  and  $p_s$ .

```
model{
  for(i in 1:n){
    x[i] ~ dbin(pr[i],K)
    pr[i] <- p[i]*ps
    p[i] ~ dunif(0,1)
  }
  ps ~ dunif(0,1)
}
```

## 1.3 Zero inflated model for microbial counts

Reference:

Aarnisalo K, Vihavainen E, Rantala L, Maijala R, Suihko M-L, Hielm S, Tuominen P, Ranta J, Raaska L. Use of results of microbiological analyses for risk-based control of *Listeria monocytogenes* in marinated broiler legs. *International Journal of Food Microbiology*, 2008, 121: 275-284.

The data set represents 165 observations of *Listeria* concentrations in broiler legs (CFU/g, transformed to counts per 10 gram). Most observations are zeros, and only a few are fairly high. Any standard distribution is unlikely to describe the data well. A zero inflated model might be reasonable.

```
model{
  # listeria in broiler legs
  for(i in 1:N){
    cfu10g[i] ~ dpois(la[i])
    la[i] <- lambda[index[i]]*(1-U[i])
    index[i] <- 1*equals(company[i],1)+
               2*equals(company[i],8)+
               3*equals(company[i],9)
    U[i] ~ dbern(pu0[i])
    pu0[i] <- p0[1]*equals(company[i],1)+
              p0[2]*equals(company[i],8)+
              p0[3]*equals(company[i],9)
  }
  log(lambda[1]) <- beta[1]
```

```

log(lambda[2]) <- beta[2]
log(lambda[3]) <- beta[3]
for(i in 1:3){beta[i] ~ dnorm(0,0.0001)
  p0[i] ~ dunif(0,1); E[i] <- lambda[i]*(1-p0[i]) }
}

list(N=165,
cfu10g=c(0,0,0,0,0,0,20,0,0,0,3,0,3,0,0,8,0,0,0,0,0,0,0,0,0,0,5,10,13,
3,0,3,28,14,5,211,5,0,0,3,18,0,0,5,8,0,10,100,0,0,0,0,0,0,0,0,0,
0,64,0,55,0,70,0,35,0,0,0,0,0,0,5,27,0,0,20,84,0,68,0,52,0,0,34,
0,0,5,0,0,148,0,1470,0,0,0,0,0,0,5,8,5,0,0,0,5,0,8,5,103,0,0,25,
0,35,0,0,0,10,8,0,0,0,13,0,0,13,0,0,18,0,0,0,5,18,8,0,27,0,9,0,0,
0,0,26,0,0,3,0,0,297,0,5,0,5,0,0,0,0,27,0,0,15,0,0,0,0,0,0),
company=c(1,9,1,1,1,1,1,9,9,9,9,9,9,1,8,8,1,8,9,1,1,8,9,8,8,8,
8,9,8,8,8,8,9,9,1,1,8,1,1,1,9,9,8,8,1,1,8,1,1,1,1,1,8,1,8,9,8,1,
8,9,8,1,1,9,9,1,1,8,8,9,9,8,8,1,8,9,8,1,8,8,9,9,8,9,9,8,9,8,9,9,
9,1,9,9,8,8,8,1,1,9,1,1,9,8,1,1,1,9,9,8,1,8,8,9,9,8,9,1,8,8,9,8,
9,1,1,9,1,1,9,8,9,8,9,1,8,9,8,1,9,8,1,1,1,1,1,1,1,9,1,1,8,9,8,8,
8,8,9,8,8,9,9,9,9))

```

## 1.4 Change point of Poisson intensity

The change point estimation of coal mine accidents in Britain, 1851-1962, can be modeled as Poisson-process with intensity  $\mu = \exp(\beta_1)$  before the change, and  $\mu = \exp(\beta_1 + \beta_2)$  after the change. The parameters, and the change point are to be estimated.

Decade	Number of accidents									
1851-1860	4	5	4	1	0	4	3	4	0	6
1861-1870	3	3	4	0	2	6	3	3	5	4
1871-1880	5	3	1	4	4	1	5	5	3	4
1881-1890	2	5	2	2	3	4	2	1	3	2
1890-1900	1	1	1	1	1	3	0	0	1	0
1901-1910	1	1	0	0	3	1	0	3	2	2
1911-1920	0	1	1	1	0	1	0	1	0	0
1921-1930	0	2	1	0	0	0	1	1	0	2
1931-1940	2	3	1	1	2	1	1	1	1	2
1941-1950	4	2	0	0	0	1	4	0	0	0
1951-1960	1	0	0	0	0	0	1	0	0	1
1961-1962	0	0								

```

model{
for(year in 1:N){
T[year] <- year+1850
D[year] ~ dpois(mu[year])
log(mu[year]) <- b[1]+step(year-changeyear)*b[2]
}
for(j in 1:2){b[j]~dnorm(0,0.0001)}

```

```

changeyear ~ dunif(1,N)
mu1 <- exp(b[1])
mu2 <- exp(b[1]+b[2])
}
list(D=c(4, 5, 4, 1, 0, 4, 3, 4, 0, 6,
3, 3, 4, 0, 2, 6, 3, 3, 5, 4,
5, 3, 1, 4, 4, 1, 5, 5, 3, 4,
2, 5, 2, 2, 3, 4, 2, 1, 3, 2,
1, 1, 1, 1, 1, 3, 0, 0, 1, 0,
1, 1, 0, 0, 3, 1, 0, 3, 2, 2,
0, 1, 1, 1, 0, 1, 0, 1, 0, 0,
0, 2, 1, 0, 0, 0, 1, 1, 0, 2,
2, 3, 1, 1, 2, 1, 1, 1, 1, 2,
4, 2, 0, 0, 0, 1, 4, 0, 0, 0,
1, 0, 0, 0, 0, 0, 1, 0, 0, 1,
0, 0),N=112)

```

## 1.5 Estimating within flock prevalence for egg laying flocks

Reference:

[http://www.fsis.usda.gov/Science/Risk\\_Assessments/index.asp](http://www.fsis.usda.gov/Science/Risk_Assessments/index.asp)

Risk Assessment for Salmonella Enteritidis in Shell Eggs and Salmonella spp. in Egg Products (Oct 2005).

In this example, it is assumed that the flocks are either completely free of salmonella (within flock prevalence zero), or that there is some positive prevalence in them. Furthermore, a false negative rate of  $h = 15\%$  is assumed for the testing. 58 pooled samples are analyzed from each flock. Each pooled sample consists of 5 individual sub-samples combined. The unknown fraction of completely clean flocks is  $p_0$ . The within flock prevalence of the  $i$ th flock, if nonzero, is  $p_i$  and  $\text{logit}(p_i) \sim N(\mu, \sigma^2)$ .

Below,  $\mathbf{x}$  is the number of observed positive pooled samples (out of 58), and  $\mathbf{nf}$  is the number of flocks with the corresponding result.

```

# US data: h=false negative rate (assumption)
list(K=32, h = 0.15, S=58, ss=5,
x=c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,
21,22,23,24,25,26,27,28,36,39,42,44),
nf=c(464,77,39,23,18,9,6,8,7,8,4, 6,4,4,2,2,6,1,3,3,
2,3,1,1,1,2,2,1,1,1,1,1))

```

Write hierarchical WinBUGS model for these data. Compute a predicted within flock prevalence from the hierarchical model.

```

model{
pzero ~ dunif(0,1) # P(flock is uninfected | pzero) = pzero
mu ~ dnorm(0,0.001)
tau ~ dgamma(0.001,0.001)

logit(ep) <- mu # ep = mean prevalence for infected flocks
ep.all <- ep*(1-pzero) # ep.all = mean prevalence for all flocks
zeronew ~ dbern(pzero) # predicted status for a new flock
# predicted within prevalence for a new flock:
pnew <- (1-zeronew)*pnew0 + zeronew*0;
logit(pnew0) <- logitpnew0; logitpnew0 ~ dnorm(mu,tau)
# pnew0= predicted within prevalence for a new flock, given it's infected

for(i in 1:K){ # loop over data
for(j in 1:nf[i]){ # loop over number of flocks
xx[i,j] <- x[i]
xx[i,j] ~ dbin(pr[i,j],S); # S = number of pooled samples per flock
pr[i,j]<-(1-zeroflock[i,j])*(1-pow(1-p[i,j],ss) )*(1-h) +zeroflock[i,j]*0
# ss = number of individual samples pooled
zeroflock[i,j] ~ dbern(pzero) # status
logit(p[i,j]) <- logitp[i,j]; logitp[i,j] ~ dnorm(mu,tau) # prevalence
}}
}
# US data: h=false negative rate (assumption)
list(K=32, h = 0.15, S=58, ss=5,
x=c(0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,
18,19,21,22,23,24,25,26,27,28,36,39,42,44),
nf=c(464,77,39,23,18,9,6,8,7,8,4, 6,4,4,2,2,6,1,3,3,
2,3,1,1,1,2,2,1,1,1,1,1))
list(mu=0,tau=1) # initial values

```

## 1.6 Multinomial model

Estimate the relative allele frequencies in a population.  $P(a) = p$ ,  $P(A) = q$ . The probabilities of individuals with genotypes  $aa, aA, AA$  are  $p^2, 2pq, q^2$  in equilibrium, if no inbreeding. In a sample of  $N$  individuals the observed frequencies were  $x_1, x_2, x_3$ . Below also a model with inbreeding coefficient.

```

model{ # no inbreeding
x[1:3] ~ dmulti(pr[],N)
pr[1] <- p*p
pr[2] <- 2*p*q
pr[3] <- q*q
p ~ dunif(0,1); q <- 1-p
}

model{ # inbreeding
x[1:3] ~ dmulti(pr[],N)
pr[1] <- p*p + f*p*q

```

```
pr[2] <- 2*(1-f)*p*q
pr[3] <- q*q + f*p*q
p ~dunif(0,1); q <- 1-p
f <- w*(1-fmin)+fmin; w ~dunif(0,1)
fmin <- max(-p/q, -q/p)
}
```